MATTHEW COLLINS

Machine Learning Engineer

+447736096669 matthewqcollins@gmail.com
/in/matthewqcollins/ github.com/mattcodeins

SUMMARY

Developer with a mathematically deep understanding of a wide range of ML algorithms, and the practical skills to implement them in production. Experience leading development. I like finding big ideas to solve big problems, and I care about using statistical methods that work.

SKILLS

Languages

Python, Golang, SQL, Matlab, Starlark, Bash **(Proficient)**
C++, C, C#, Java, JavaScript, Elm **(Basic)**

Technologies

ML libraries (e.g. NumPy, Pandas, PyTorch, Scikit-learn, LangChain), Git, Docker, Elasticsearch, Postgres, Bazel

EXPERIENCE

1/2023 - now **Machine Learning/Software Engineer** **Next DLP** - ***data loss and insider risk saas solution startup***

- Member of the Insights team, responsible for using (and expanding) our software's extensive data tracking agent to **develop models** for identifying anomalous or risky activity, and shipping those models to clients in a robust, well-defined, and easily-configurable manner through our online platform.
- Independently created a novel real-time user-level **anomaly detection machine learning** (meta-)model for tracking file transfers to external storage devices, using extensive analysis on customer data.
- Developed an **AI assistant** to summarise and contextualize sequences of behaviour detected as high-risk via LLMs, integrating our database into prompt chains to create informative and concise responses.
- Wide variety of merge requests: Frontend (Elm, APIs), Backend (Golang, C++, C), Plug-ins/Extensions (JavaScript, C#), Databases (Elasticsearch, Postgres), and our Model Microservice (Starlark, Golang).
- Founded and led an Artificial Intelligence paper reading group.

EDUCATION

9/2021 - 9/2022 **MSc Artificial Intelligence - *Merit*** **Imperial College London**

- Averaged **80%** in taught modules, and **66%** in individual project *(grade is minimum of these)*.
- Awarded *The Corporate Partnership Programme MSc Group Project Prize*.
- **Modules** (**First** in all): Python, Machine Learning, Deep Learning, Reinforcement Learning, Natural Language Processing, Computer Vision, Probabilistic Inference, Symbolic AI, and AI Law and Ethics.
- Implemented: Deep/Convolutional/Recurrent/Bayesian Neural Networks, Transformers, Variational Autoencoders, GANs, Gaussian Processes, Monte Carlo/Temporal Difference Methods, Random Forests, etc.

9/2018 - 6/2021 **BSc Mathematics - *First*** **University of Warwick**

- **3rd Year: 72.2% -** *Relevant Modules*: Mathematics of Machine Learning, Probability Theory, Multivariate Statistics, Matrix Analysis & Algorithms, Computational Modeling of Partial Differential Equations.
- **1st/2nd Year: 79.1/79.4% -** *Modules in topics from*: Linear Algebra, Analysis, Probability, Statistics, Geometry, Calculus, Number Theory, Discrete Mathematics, and Programming (Matlab and Java).

9/2011 - 6/2018 **A Levels - *A*A*A**** *(Maths, Further Maths, Physics)* **Bishop Wordsworth's Grammar School**

- The 6th Form Award for Mathematics (Speech Day 2018) • Prefect • Gold D of E.

SELECTED PROJECTS

1/2022 - 4/2022 **MSc Group Project - Tracking Brand Sentiment via Public News** **github link**

- Leader of a 6-person group working with corporate partner, theTradeDesk, to develop a daily brand sentiment analysis pipeline and display, powered by online news ($\sim$ 1 million scraped articles/day).
- Automated **AWS** pipeline (EMR, EC2, S3, Batch, Lambda, Glue, Athena, QuickSight), incorporating our Named Entity Recognition and Sentiment Classification models using Hugging Face and Apache Spark.

5/2022 - 9/2022 **MSc Individual Project** **github link**

- Research into Marginal Likelihood based Model Selection for **Bayesian Deep Learning**.
- Extended **PyTorch** to allow implementation of Bayesian Neural Networks (with theoretical bounds).

2022 **Hackathons**

- **Imperial College Hack** - Web-app for locating optimal areas for solar panels using satellite imaging via Google Cloud Platform and deep-learning terrain recognition. **github link**
- **Google Hash Code** - Global competition to simulate and maximise a real-world intractable optimisation problem in 4 hours. 255th out of 10,000 competing, and 1st from Imperial. **github link**
- **CarbON/CarbOFF** - CO_2e forecasting app for households. 2nd prize & best presentation. **github link**

HOBBIES

Climbing/Bouldering, Board Games, Cycling, Philosophy.